

代数 1 • 单位 3 工作表

二变量统计

姓名： _____ 日期： _____

A 部分 • 双向表

1. 一项针对 200 学生的调查询问他们更喜欢狗还是猫，以及他们住在房子还是公寓里。
- a. 住独立屋的人中，喜欢狗的占几分之几？ _____
 - b. 爱狗人士分数住在什么房子里？ _____
 - c. 为什么 (a) 和 (b) 不同，尽管两者都涉及“狗”和“房子”？用你自己的话解释一下。

d. 住房类型和宠物偏好之间是否存在关联？使用相对频率证明合理性。

	狗	猫	总计
房子	84	36	120
公寓	32	48	80
总计	116	84	200

B 部分 • 散点图和线性模型

2. 散点图显示 12 学生的鞋码 (x) 与身高（以英寸为单位）(y)。最佳拟合线是 $y = 1.4x + 50$ 。
- a. 斜率 1.4 在上下文中意味着什么？（使用单位。）

b. y-截取 50 是什么意思？这种解释在这里合理吗？解释。

c. 用鞋码 11 预测人的身高。_____ 英寸

d. 用鞋码 25 预测人的身高。你相信这个预测吗？为什么或为什么 not 呢？

C 部分 · 残差和拟合优度

3. 对于数据集，线性模型为 $\hat{y} = 2x + 5$ 。

a. 观察结果是 (3, 13)。计算残差。剩余 = _____

b. 另一个观察结果是 (7, 17)。计算残差。剩余 = _____

c. 如果残差为正或负，这意味着什么？

d. 如果大多数残差聚集在零附近，这对模型意味着什么？

4. 两张散点图均具有相关性 系数 $r \approx 0.85$ ，但一张散点图具有明显的弯曲图案，而另一张散点图看起来与随机散点呈线性关系。

线性模型可以更好地描述哪一个？残差分析如何帮助您做出决定？

★这是一个经典的警告：高 $|r|$ 单独并不能保证良好的线性拟合。

D部分 · 相关系数

5. 将每个 散点图 描述与可能的相关性 系数 进行匹配。选择： $r \approx -0.95$, $r \approx -0.3$, $r \approx 0$, $r \approx 0.4$, $r \approx 0.97$

- a. 强劲、近乎完美的上升趋势。 $r \approx$ _____
- b. 强劲的下降趋势，分散程度较小。 $r \approx$ _____
- c. 无清晰图案；点看起来是随机的。 $r \approx$ _____
- d. 弱势积极趋势；很多分散的。 $r \approx$ _____
- e. 弱势负面趋势；很多分散的。 $r \approx$ _____

E 部分 · 相关性与因果性

6. 一名研究人员发现，随着城市冰淇淋销量的增加，溺水死亡人数也会增加。相关性很强且呈正相关。

a. 冰淇淋会导致溺水吗？为什么这两者可能相关呢？

b. 说出可以解释这两种趋势的第三个变量。

7. 对于每一对，确定这种关系是否最有可能是因果关系、仅仅是相关（共同原因）还是巧合。简要说明理由。

a. 学习与考试成绩。

b. 火灾中的消防员人数与造成的损失。

c. 全球海盗数量与全球平均气温。

d. 每日卡路里摄入量与体重增加。

F · 误差分析与综合

8. 林说：“我朋友的鞋码和数学考试成绩之间的相关性是 0.4。这意味着脚越大，数学成绩越好。”

a. 林先生的说法有什么问题吗？确定至少两个问题。

b. 提出更合理的解释。

9. 一项研究发现两个变量之间存在 $r = 0.99$ 。

a. 我们应该自动使用线性模型来预测吗？为什么或为什么不呢？

b. 在信任该模型之前，您还会做哪些其他检查？

10. 打开：设计你自己的研究。

a. 确定您认为可能相关的两个数值变量：

b. 您预测 r 的什么征象？强还是弱？

c. 是否存在令人困惑的（第三个）变量？说出一个。

d. 如果您发现了 $r = 0.6$ ，您能否声称其中一个原因是另一个原因？为什么或为什么不是呢？
